

Health State Survey-Derived Utilities in Cost-Utility Analysis—A Call to Action

Benjamin P. Geisler, MD, MPH^{1,2,3}

¹Institute for Technology Assessment, Massachusetts General Hospital/Harvard Medical School, Boston, MA, USA; ²Division of Oncology, Beth Israel Deaconess Medical Center/Harvard Medical School, Boston, MA, USA; ³Department of Public Health, Medical Decision Making and Health Technology Assessment, UMIT, Hall i.T., Austria

Value in Health has recently devoted a special issue to the quality-adjusted life years (QALY) (Volume 12, Issue s2), essentially depicting it as the only realistic alternative to value health outcomes. Although some of the QALYs' shortcomings were described, the focus of the special issue was on "building a pragmatic road," developing avenues to possible solutions for the problems at hand.

In the current issue, Joore et al. raise yet another, potentially serious flaw for the application of QALYs in cost-utility analysis [1]. The authors used five dataset of mostly piggy-back economic analyses, where both the widely used health state surveys EQ-5D and SF-6D (a SF-36 derivative) health utility state surveys had been conducted, to compute cost-effectiveness acceptability curves via probabilistic sensitivity analysis. The main finding of their elegant study is that the small differences in utilities that were derived from the two different surveys might result in substantially different results when applied in cost-utility analysis. This potentially changes the interpretation of the results of cost-utility analyses which has implications for the decision-makers.

Although a Medline search reveals around 80 articles which both mention EQ-5D and SF-6D, the earliest one dating back to 2001, the differences in EQ-5D- or SF-6D-derived utilities are at present still unclear. Ceiling and floor effects as well as different sensitivities in milder health states have been described a while ago; it was also shown that different scoring in the survey instrument with economic valuation methods such as time trade-off, standard gamble, or visual analog scale might produce significantly different utilities. Nevertheless, the impact on incremental QALYs and incremental cost-utility ratios is still unclear. The present study adds that SF-6D-derived utilities, despite being scored by a time trade-off-valuated algorithm, are not always higher and might not always have narrower ranges. The EQ-5D, on the other hand, might not always be less sensitive in milder health states. Finally, the choice of the instrument can in some cases flip the decision, from inferior to cost-effective, or from dominant to not cost-effective.

The authors conclude that "a systematic difference in the probability of accepting the cost-utility of interventions as a result of the choice of utility instrument would seriously bias the comparability of the results of economic evaluations." Nevertheless, not only systematic but also occasional differences could seriously

damage the reputation of the QALY as a universal "currency" of health benefits across conditions and intervention strategies.

There are three implications from this study. First, there needs to be more research on the consequences of using either EQ-5D or SF-6D. If the five studies chosen by Joore et al., and this is likely, are not the only ones affected by the choice of the survey instrument, we need guidance from ISPOR for practice. Second, EQ-5D- and SF-6D-derived utilities are valued by two different methods, time trade-off and standard gamble, which are considered the gold standard from an economics perspective. The decision-analytic community should consider standardizing on either one instrument or a common rescaling on both instruments. Third, the results from Joore et al. point at a potential for misuse. Several of their cost-effectiveness acceptability curves slope upward or downward depending on the survey instrument used. Manufacturers submitting decision-analytic models for appraisal processes to, for example, the U.K.'s National Institute for Clinical Excellence could "cherry-pick" the more favorable result. Moreover, the existence of the other, less favorable utility dataset could be obscured as in recent controversies around clinical efficacy datasets. ISPOR should consider recommending mandatory prospective registration of health state assessments, similarly to the U.S. Food and Drug Administration's ClinicalTrials.gov or the Cochrane Central Register of Controlled Trials. Economic piggy-back analyses could even be registered in these databases. It is up to us to act upon these alarming findings, and preserve the credibility of the QALY.

Acknowledgment

The author would like to thank Dennis G. Fryback, PhD, for his helpful comments.

Source of financial support: None.

Reference

- 1 Joore M, Brunenberg D, Nelemans P, et al. The impact of differences in EQ-5D and SF-6D utility scores on the acceptability of cost-utility ratios: results across five trial based cost-utility studies. *Value Health* 2009 doi: 10.1111/j.1524-4733.2009.00669.x.

Address correspondence to: Benjamin P. Geisler, Beth Israel Deaconess Medical Center/Massachusetts General Hospital, 38 Calvin Street, Somerville, MA 02143, USA. E-mail: ben.geisler@gmail.com
10.1111/j.1524-4733.2009.00665.x